A

Topic Approval Seminar

on

Robust Multimodal Techniques for Comprehensive Mental Health Analysis

Submitted in partial fulfilment of the requirement for the degree of PhD

In

Science and Technology

(Computer Engineering)

By

Rupali Yashwant Landge
Admission/ Registration No: 2220220443

Under the guidance of
Dr. Hariram Chavan

Department of Computer Engineering
K J Somaiya Institute of Technology
Somaiya Ayurvihar Complex Eastern Express Highway Near Everard Nagar, Sion East, Mumbai, Maharashtra 400022

मुंबई विद्यापीठ
University of Mumbai

2022-23

Faculty of Technology

PROPOSAL FOR TOPIC APPROVAL TO BE SUBMITTED TO THE UNIVRSITY OF MUMBAI FOR THE PH.D. DEGREE IN THE COMPUTER ENGINEERING OF SCIENCE AND TECHNOLOGY

Title of Proposal	:	Robust Multimodal Techniques for Comprehensive Mental Health Analysis
Name of the Student	:	Rupali Yashwant Landge
Basic Qualification	:	M.Tech CSE
In the case of Ph. D. students Year of Passing of PET/NET/SET/M. Phil. / GPAT / GATE (Teaching Experience)	:	PET, August 2022
Name & Designation of the Research Supervisor / Guide	:	Dr. Hariram Chavan
Date of Superannuation/Retirement of the Research Supervisor/Guide		01-06-2032
Name of the Research Centre	:	K. J. Somaiya Institute of Technology, Sion, Mumbai.
Signature of Student	:	
Signature of Research Supervisor / Guide	:	
Table of Contents	:	

Contents	Page No.
1. Introduction	1
2. Literature Survey	6
3. Objectives	18
4. Methodology	19
5. Scheduling and Planning	21
6. Conclusion	22
References	23

Date: ~~08-11-2023~~ 10/11/2023

Signature of Head / Director / Principal Research Centre

Principal
K J Somaiya Institute of Technology
Sion (E), Mumbai - 400 022

K J Somaiya Institute of Technology Sion Mumbai 400 022

CERTIFICATE OF APPROVAL

This is to certify that Rupali Yashwant Landge has delivered a Topic approval seminar on **"Robust Multimodal Techniques for Comprehensive Mental Health Analysis"** on 31/8/2023 and submitted a report in **K J Somaiya Institute of Technology** Sion Mumbai, in the partial fulfilment of the degree of PhD in the Science and Technology (**Computer Engineering**).

Dr. Hariram Chavan

Guide

Dr.Sunita Patil

Subject Expert Internal

Dr.Mahesh Shirole

Subject Expert External

Dr.Sarita Ambadekar

Head of the Department

Dr. Hariram Chavan

PhD Coordinator

Dr. Vivek Sunnapwar

Principal

KJSIT,Sion

Principal
K J Somaiya Institute of Technology
Sion (E), Mumbai - 400 022

K J Somaiya Institute of Technology
Sion
Mumbai
400 022

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Rupali Yashwant Landge

Registration No:2220220443

Date:

Place: Sion

ACKNOWLEDGEMENT

A few sublime human experiences defy expressions of any kind, and a feeling of true gratitude is one of them. I, therefore, find words quite inadequate to express my indebtedness to my Guide and PhD coordinator **Dr. Hariram Chavan**, his virtuous guidance, encouragement and help throughout this work. Their deep insight into the problem and the ability to provide solutions has been of immense value in improving the quality of work at all stages. This experience of working with them shall ever remain a source of inspiration and encouragement for me.

I express my thanks to **Dr. Sunita Patil, Vice Principal KJSIT**, Sion for extending her support that she gave to truly help the progression of the topic approval seminar work.

My sincere thanks to **Dr.Vivek Sunnapwar Principal, KJSIT**, Sion for providing us the necessary assistance in the completion of the work.

I am extremely grateful to the celebrated authors whose precious works have been consulted and referred to in my work. I also wish to convey my appreciation to my friends who provided encouragement and timely support in the hour of need.

Special thanks to my Parents and family whose love and affectionate blessings have been a constant source of inspiration in making this a reality.

All the thanks are, however, only fraction of what is due to Almighty for granting me an opportunity and the divine grace to successfully accomplish this assignment

Rupali Yashwant Landge

Registration No: 2220220443

ABSTRACT

There is a global rise in mental health problems that impact not just adults but also teenagers under the age of 18 and children. Undoubtedly, a person's emotions, intellect, and ability to communicate with others are all affected by mental illness. Mental health disorders increase the risk for other diseases and contribute to unintentional and intentional injury. Mental health problems affect the entire society, not just a small, isolated portion of it. These issues have shown that mental illness gives serious consequences across societies and demands new strategies for prevention and intervention. Early detection is essential in order to better understand the patient's health situation and give them better care because many mental health issues have the potential to be lethal.

The proposed system mainly focuses on the integration of text, audio and video data streams to record a variety of behavioural and emotional indicators. Video features cover facial expressions, body language, and eye movement patterns while audio features cover speech prosody, sentiment analysis, and sonic characteristics and text based features cover sentimental analysis. For feature extraction model can use machine learning and deep learning techniques. A multimodal network integrates the collected features and merges the text, audio and visual data enables a thorough examination of a person's mental health. The combination of these modalities provides more effective mental health intervention and support programmes in the modern digital era by enabling a more nuanced and accurate assessment of a person's mental state.

TABLE of CONTENT

Acknowledgment			i
Abstract			ii
Table of Contents			iii
List of Figure			iv
List of Table			iv
1	Introduction		1
	1.1	Different types of mental illness	2
	1.2	About multimodal model	4
	1.3	Need of multimodal model in mental health analysis	4
	1.4	Aim of mental health analysis	5
2	Literature Review		6
	2.1	Challenges and gaps in multimodal mental health analysis	17
3	Objectives of multimodal mental health analysis model		18
4	Methodology for multimodal mental health analysis model		19
5	Scheduling and Planning		21
6	Conclusion		22
7	References		23

LIST OF TABLES

Table No.	Name	Page no
Table 2.1	Summary of Literature	17
Table 5.1	Scheduling and Planning	21

LIST OF FIGURES

Figure No.	Name	Page no
Fig 4.1	Proposed multimodal mental health analysis model	19

Chapter 1 Introduction

Mental health is a condition of mental wellness that helps people to manage life's stresses, realise their potential, learn and work effectively, and contribute back to their communities. It is a crucial element of health and well-being that supports both our individual and group capacity to decide, form connections, and influence the world we live in. Mental health is a basic human right and it is crucial to personal, community and socio-economic development. Mental health is more than the absence of mental disorders. It is experienced differently from person to person along a complex continuum, with varied levels of difficulty and distress as well as potentially very different social and clinical effects. A mental disorder is characterized by a clinically significant disturbance in an individual's cognition, emotional regulation, or behaviour. It is usually associated with distress or impairment in important areas of functioning. Mental disorders may also be referred to as mental health conditions.

Around 450 million people globally have mental illness problems. It is expected to affect at least one member of one in every four families (WHO, 2001). The harmful effects of stigma and prejudice are also carried by these families in addition to offering them physical and emotional support. Over 90% of them are still untreated. The huge treatment gap is mostly due to a lack of understanding about the signs of mental illness, stigma and myths associated with it, as well as ignorance about the treatments that are available and the possible advantages of getting therapy. Mental health illnesses and sickness that not onlyaffect children but also adolescents under the age of 18 and adults.Early detection is essential in order to better understand the patient's health situation and give them better care because many mental health issues have the potential to be lethal. Instead of depending on any lab tests, the diagnosis of these mental health illnesses is based on self-report and linked health questionnaires that are specifically created to detect certain sensations or interactions. Artificial intelligence and machine learning technologies are increasingly being used to assist mental health providers in better understanding the mental conditions, thereby assisting in improved decision making, as the amount of data available for such studies related to mental health is growing exponentially. One of the most recent developments in artificial intelligence is deep learning, which imitates how the human brain processes data and has the

capacity to learn, recognise, and discover patterns. This technology aids in decision-making by allowing systems to recognise patterns in data and learn new ones.

1.1 Different types of Mental Illness

There are many different conditions that are recognized as mental illnesses. The more common types discuss below.

- **Depression:** A medical disorder, depression is one of the most prevalent mental illnesses that affect millions of individuals worldwide. Depression is seen as a serious illness that can be physically harmful to a sufferer as well as affect how they feel mentally. According to a patient's mental health status, depression severity can be predicted.

  Depression is a form of mental disorder where the sufferer experiences persistent feelings of helplessness, demotivation, mood fluctuations, and loss of interest in daily physical, mental, and social activities. This result in emotional harm and physical changes to the patient's body. It specifically impacts a person's ability for learning, produces mood swings, and frequently lowers a person's level of work efficiency. Depending on its degree, depression can cause a variety of symptoms in a patient. Furthermore, it can have a severe impact on a person's thoughts, behaviour's, work-life, and eating habits. Usually, the depressed person may feel sadness, anxiety, hopeless, worried, irritated, or restless. Brain activity slows down and the hormone cortisol is produced in high severity, which has an impact on the development of new neurons in the brain. A person's mental process is severely affected, and it can occasionally result in suicidal circumstances. The World Health Organization (WHO) ranks depression as one of the top 10 contributors to the global disease burden and predicts it will be the leading contributor by 2030[1].

- **Anxiety Disorder**: Anxiety is a feeling of fear, dread, and uneasiness. It might cause you to sweat, feel restless and tense, and have a rapid heartbeat. It can be a normal reaction to stress. For example, you might feel anxious when faced with a difficult problem at work, before taking a test, or before making an important decision. It can help you to cope. The anxiety may give you a boost of energy or help you focus. But for people with anxiety disorders, the fear is not temporary and can be overwhelming.

Anxiety disorders are conditions in which you have anxiety that does not go away and can get worse over time. The symptoms can interfere with daily activities such as job performance, schoolwork, and relationships.

- **Post-traumatic stress disorder (PTSD)**: PTSD is a disorder that develops in some people who have experienced a shocking, scary, or dangerous event. It is natural to feel afraid during and after a traumatic situation. Fear is a part of the body's "fight-or-flight" response, which helps us avoid or respond to potential danger. People may experience a range of reactions after trauma, and most people recover from initial symptoms over time. Those who continue to experience problems may be diagnosed with PTSD.
- **Bipolar Disorder:** People with bipolar disorder experience alternating depressive episodes with periods of manic symptoms. During a depressive episode, the person experiences depressed mood (feeling sad, irritable, empty) or a loss of pleasure or interest in activities, for most of the day, nearly every day. Manic symptoms may include euphoria or irritability, increased activity or energy, and other symptoms such as increased talkativeness, racing thoughts, increased self-esteem, decreased need for sleep, distractibility, and impulsive reckless behaviour. People with bipolar disorder are at an increased risk of suicide. Yet effective treatment options exist including psychoeducation, reduction of stress and strengthening of social functioning, and medication.
- **Schizophrenia**: Schizophrenia is characterised by significant impairments in perception and changes in behaviour. Symptoms may include persistent delusions, hallucinations, disorganised thinking, highly disorganised behaviour, or extreme agitation. People with schizophrenia may experience persistent difficulties with their cognitive functioning.
- **Eating disorders**: Eating disorders such as anorexia nervosa and bulimia nervosa, involve abnormal eating and preoccupation with food as well as prominent body weight and shape concerns. The symptoms or behaviours result in significant risk or damage to health, significant distress, or significant impairment of functioning. Anorexia nervosa often has its onset during adolescence or early adulthood and is associated with premature death due to medical complications or suicide.

Depression, Anxiety disorders and Post-traumatic stress disorder (PTSD) these three conditions make up around 30 percent of all diagnoses of mental illness. While they share many of the same qualities, they're also significantly different from one another.

1.2 About Multi-Modal Model

A multimodal model is an artificial intelligence (AI) system designed to process several sensory inputs simultaneously, much to how people perceive their surroundings. Different from normal unimodal AI systems, which are trained to do a certain job using a single sample of data, multimodal models are trained to integrate and analyse data from several sources, including text, images, audio, and video. This method incorporates the context and supporting data necessary to generate accurate projections, giving a more comprehensive and nuanced picture of the data. In contrast to unimodal systems, multimodal models produce more dynamic predictions by combining diverse data from several sensors and data inputs into a single model.

1.3 Need of Multimodal Model in Mental Health Analysis

Following are some ways how multimodal model used in mental health analysis

1. **Emotion Recognition:** Multimodal models can be used to analyse a person's facial expressions, speech patterns, and text content to accurately identify and quantify emotional states. This can assist in monitoring mood changes and detecting potential signs of depression, anxiety, or other mood disorders.
2. **Sentiment Analysis:** By combining text and visual data from social media posts, messages, and images, multimodal models can provide insights into an individual's overall sentiment and emotional state over time. This can help identify trends in emotional well-being and potential shifts towards negative emotions.
3. **Early Detection of Mental Health Issues:** Multimodal models can analyze a variety of data sources, such as speech tone, facial expressions, and language patterns, to detect early signs of mental health issues. This could lead to timely interventions and support, preventing more serious conditions from developing.
4. **Personalized Treatment Plans:** By integrating multimodal data from various sources, including self-reported information, sensor data, and video consultations,

these models can assist mental health professionals in tailoring personalized treatment plans that address an individual's specific needs and challenges.

5. **Virtual Support Systems:** Multimodal models can be integrated into virtual support systems, chatbots, or digital therapists. These systems can provide continuous, real-time support by analysing various cues like speech, text, and user interactions to offer coping strategies, resources, and encouragement.
6. **Remote Monitoring:** Multimodal models allow for remote monitoring of mental health patients, providing clinicians with a holistic view of the patient's well-being. This is especially valuable for individuals who have limited access to in-person care.
7. **Suicide and Crisis Prevention:** By analysing text and other cues, multimodal models can help identify individuals who might be at risk of self-harm or suicide. This can enable timely interventions and connect them with appropriate support resources.
8. **Therapeutic Interventions:** In therapeutic settings, multimodal models can assist therapists by analysing the patient's verbal and nonverbal cues, helping them gain deeper insights into the patient's emotional state and progress during sessions.
9. **Research Insights:** Researchers can use multimodal models to analyze large datasets from multiple sources, helping to uncover patterns, correlations, and risk factors associated with different mental health conditions.
10. **Reducing Stigma:** Using multimodal models for mental health analysis can provide an objective and data-driven approach to understanding mental well-being, potentially reducing the stigma associated with seeking help for mental health issues.

1.4 Aim of Mental health analysis

The aim of mental health analysis is to assess and understand an individual's mental and emotional well-being to identify any potential mental health issues or conditions. Mental health analysis is to promote early intervention, accurate diagnosis, personalized treatment, and improved mental health outcomes for individuals experiencing mental health challenges.

Chapter 2 Literature Review

Currently Mental health is a serious issue in societies. Considerable advancement in the field of mental health analysis is required to address this issue. Researchers, doctors have been compelled to delve more deeply into the field of mental health analysis because of the complex interplay of psychological, biological, and social elements in determining mental well-being. This survey gives insights about work already done in this domain, challenges need to be addressed and future research opportunities available.

Ming Fang, Siyu Peng , Yujia Liang , Chih-Cheng Hung , Shuhua Liu thoroughly examine audio-visual and textual data on depression in order to fully extract its properties. We next present a multimodal fusion model with multi-level attention mechanism (MFM-Att) for depression identification. The process can be broken down into two steps, the first of which uses two LSTMs and a Bi-LSTM with an attention mechanism to learn multi-view audio, visual, and rich text features, respectively. The attention fusion network (AttFN), which aims to utilise the diversity and complementarity between modalities for depression detection, is fed the output features of the three modalities in the second stage to acquire useful depression information. It is important to note that the multi-level attention mechanism may learn the correlations of inter-modality in addition to extracting useful depressive aspects from intra-modality, which enhances the model's overall performance by lowering the impact of duplicated input [1].

A multi-modal depression detection technique based on high-order emotional characteristics was proposed by Yuran Ru,Ning Geng,Li Li. A two-stage network is created to simultaneously recognise emotions and diagnose depression, with the emotional outcomes being input as high-order semantic features into the enhanced TBJE-E multimodal network. With the aid of the co-attention module, this procedure guided the learning of additional modalities and ultimately produced the predicted effects. The F1 performance of the TBJE-E model with emotional aspects is comparatively 6.3% better than that of the original TBJE model. In the depression detection test, the approach presented in this work has achieved the SOTA level. The experimental results also demonstrate that there is currently very little chance that this technology will violate an individual's internal psychological privacy without

that person's knowledge, and that this technology may be useful in other professional fields such as criminal investigation, psychological diagnosis and treatment [2].

Convolutional neural network-based deep learning techniques have displayed outstanding effectiveness in depression analysis. However, these approaches face the following significant difficulties that must be overcome: (1) Due to the spatial locality, it is still challenging for CNNs to learn long-range inductive biases in the low-level feature extraction of various facial areas. (2) A model with a single attention head finds it challenging to focus on several facial features at once, which reduces sensitivity to other significant facial areas linked to depression. Many of the cues used to identify facial depressions come from several facial features at once, such as the mouth and eyes. They provide a two-stage, end-to-end integrated structure dubbed the Hybrid Multi-head Cross Attention Network (HMHN) to address these problems. Grid-Wise Attention (GWA) and Deep Feature Fusion (DFF) blocks are used in the initial stage to learn low-level visual depression features. In the second stage, we encode high-order interactions between local features using the Multi-head Cross Attention block (MAB) and Attention Fusion block (AFB) to produce the global representation [3].

The author Hyewon Yoo and Hayoung Oh analyses the effectiveness of previous research on multimodal emotion detection and suggests a model that detects depression by fusing two modalities with input values from the speaker's text and voice. Voice characteristics were extracted using CNN, text features were extracted using Transformers, and two modalities were combined using a tensor fusion network based on the DAIC-WOZ dataset. In the final layer, we also create a model that uses LSTM to determine whether the speaker is depressed or not [4].

Vandanaa, Nikhil Marriwalab, Deepti Chaudharyb proposed a method which consists of three parts: a textual CNN model, which trains a CNN model using only text features; an audio CNN model, which trains a CNN model using only audio features; and a hybrid model, which combines an audio and textual model and uses LSTM algorithms. The suggested work additionally makes use of the Bi-LSTM model, an enhanced variation of the LSTM model. For the evaluation of models, certain parameters are found, including precision, F1-score, recall, and support. For training loss, validation loss, training accuracy, and validation accuracy, graphs are plotted in the results. Finally, depression for textual CNN Model, audio

CNN Model, LSTM model, and Bi-LSTM against true label and predicted label can be recognised using confusion matrix [7].

Irune Zubiaga and Raquel Justo attempt to establish a conversational monitoring system for emotional health. The objective of their study is to choose the features that can be useful for training a model that seeks to determine whether a patient has a mental disorder as a first step in attaining this goal. To do that, they generated a data set using transcriptions that were taken from conversational data collected from individuals with various mental health issues. The created data set will then be supplied to supervised learning algorithms, and their performance will be assessed after the feature selection. Specifically, they used BERT, neural networks, and random forests[21].

By analysing posts and comments made by a person on social networking sites, Tanya Nijhawa, Girija Attigeri and T. Ananthakrishna want to expand sentiment and emotion analysis to identify that person's level of stress. They used massive datasets of tweets to perform sentiment analysis using machine learning methods and the deep learning model BERT for sentiment classification. They also used Latent Dirichlet Allocation, an unsupervised machine learning technique for scanning a collection of documents, identifying word and phrase patterns within them, and assembling word groups and similar expressions that most accurately portray a collection of documents. These algorithms allowed to identify people' emotions when they are online. These feelings can also be utilised to examine stress or sadness. A BERT model and machine learning (ML) models both have very high detection rates. For one's mental health to be in good shape, this research is beneficial. The outcomes, which are assessed using different metrics at the macro and micro levels, show that the trained model can identify the emotional status based on social interactions [23].

Anu Priya , Shruti Garg , Neha Prerna Tigga identified five distinct severity levels of stress, depression, and anxiety using machine learning algorithms. A standard questionnaire (DASS-21) measuring the common symptoms of stress, anxiety, and depression was used to gather data. Then, five distinct classification methods were used: K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest Tree (RFT), Naïve Bayes, and Support Vector Machine (SVM). While Random Forest was determined to be the best model, naïve Bayes had the highest accuracy. The f1 score, which is utilized in situations of imbalanced partitioning, was the basis for selecting the best model because this problem resulted in imbalanced classes [24].

Elderly depression is a pervasive yet taboo problem. Tianruo Xu offers a method for identifying depression in the senior population using multimodal techniques. The Multimodal Transformer Elder Depression Detection model illustrates the conceptual foundation for the solution. The model processes the input with many types of data, such as text, audio, and video, and gathers data for data fusion. The model can generate a score that reflects the patient's mental state after data fusion. The model uses the DAIC and extended DAIC databases for testing and training. On the basis of the data acquired from the senior members of the nearby hospital, we have also established extra tests. The evaluation outcomes show that the model can accurately identify depression in elder groups [25].

By thoroughly correlating the multidimensional patient data, artificial intelligence technology may effectively aid in the assessment of mental health issues, generating solutions for the screening of common psychological issues in huge populations of college students. Therefore, in order to facilitate the widespread and normalised screening of mental health issues in colleges and universities,Haibo Xu,Xiang Wu ,Xin Liu suggests a method for assessing mental health that combines conventional scales and multimodal intelligent identification technologies. The first way is to perform health questionnaires using conventional techniques while utilising psychological assessment scales based on human-computer interaction. Second, incorporate machine learning technology to determine college students' conditions and gauge the seriousness of psychological issues. The experiments also demonstrated that the suggested multimodal intelligent recognition approach can more accurately proofread findings from normal scale tests and has a high level of accuracy.This research has some limitations. First, there are not enough samples to satisfy the requirements, and some statistical theory-based computations might be biased. Despite the fact that gender and age were evenly distributed, sampling bias may have also been influenced by the participants' common experiences. Secondly, there is a need to enhance the multimodal emotion recognition model's accuracy and applicability [26].

Ravi Prasad Thati ,Abhishek Singh Dhadwal, Praveen Kumar1 ,Sainaba P compiled a dataset of genuine people, both depressed and not. They tested different multi-modality features, feature selection methods, fused features, and machine learning classifiers for classification such Logistic Regression, Support Vector Machines (SVM), etc. The highest classification accuracy is attained when features from all three data modalities are combined, according to their findings, which show that fusing information from several modalities performs better than any single data modality. Comparing the accuracy of the feature selection approach to

other methods, Pearson's correlation coefficients significantly increased the accuracy. SVM also produced the best accuracy, at 86%. There suggested strategy was also applied to a benchmarking dataset, and the findings showed that the multimodal approach outperformed modern depression identification methods in terms of performance .The analysis of the results may be impacted by the small number of participants due to the selection criteria[27].

Using Convolutional Bidirectional LSTM as our foundation, Ping-Cheng Wei, Kunyu Peng, Alina Roitberg, Kailun Yang, Jiaming Zhang and Rainer Stiefelhagen take on the challenge of automatically recognising sadness from multi-modal data and develop a sub-attention technique for tying together disparate data. They undertake extensive trials using the publicly available DAIC-WOZ benchmark for depression assessment using various evaluation modalities and accounting for gender-specific biases in order to validate this hypothesis. The suggested model successfully detects major depression with 0.89 precision and 0.70 F1-score and successfully estimates its severity with 4.92 MAE. While learning to diagnose the condition from beginning to end and depending on significantly fewer preprocessing steps, attention-based fusion module routinely outperforms traditional late fusion approaches and achieves competitive performance in comparison to the previously reported depression estimate frameworks [29].

Methods of automatic depression recognition based on speech signals are continuously being suggested and updated for better prediction and diagnosis by Wu, Pingping; Wang, Ruihao; Lin, Han; Zhang, Fanlong; Tu, Juan; Sun, Miao, moving away from the early traditional methods based on hand-crafted features and towards the use of deep learning architectures. The most significant and recent research on automatic depression recognition by intelligent speech signal processing is presented in study in a methodical and precise manner. Additionally, techniques for acoustic feature extraction, classification and regression algorithms, and end-to-end deep models are looked at and analysed. Finally, broad trends are outlined and significant open questions are noted for future research on automatic speech depression recognition. Challenges in mental health analysis models are Availability and limitations of the baseline dataset,Model generalization,Unknown underlying correlation mechanism of acoustic informationa and Multi-modality fusion mechanism. [30]

Merging neuroimaging, neuromodulation, and cutting-edge mobile technologies in psychiatric practice, Zhe Sage Chen, Prathamesh (Param) Kulkarni, Isaac R. Galatzer-Levy, Benedetta Bigio, Carla Nasca and Yu Zhang offer a thorough assessment of ML approaches

and applications in the study.They also go over the use of ML in cross-species biomarker identification and molecular phenotyping in precision psychiatry and also address closed human-in-the-loop neuromodulation and explainable AI (XAI), emphasizing the promise of machine learning in multi-media information extraction and multi-modal data fusion. Lastly, They identify ML prospects for future research and talk about conceptual and practical issues in precision psychiatry. They discuss challenges such as, each mental disorder has overlapping symptoms with varying degrees, many disorders are presented as a spectrum, same mental disorder may have different causes and symptoms in different age/gender/race populations.[31]

Donglei Song, Qiang Shen, Haotian Feng , Rui Song , Fausto Giunchiglia , Hao Xu,presented a methodology that aims to address the difficulties in the task of detecting mental health issues. In comparison to deep learning and traditional machine learning techniques, it may firstly enhance the performance of the mental health detection task.Secondly, the explanation can be produced by the adaptive integration of spatial and temporal characteristics.The model is limited, nevertheless, by the variability of the multimodal sensory data in real life. Individuals have distinct personality traits and behavioral patterns, and depending on their mental health status, different people may exhibit different behavior aspects detected by cellphones. To effectively detect the cross-individual discrepancy, an adaptive machine learning model is therefore critically required.[5]

Proposed the use of gamification, a new trend that involves influencing user behaviour and lifestyle by adding entertaining and engaging game components to non-game contexts (such tasks relating to one's health, for example). They intend to carry out a concurrent study in which we will record videos of our subjects performing the Stroop Colour Word and Amygdala Test and analyse the footage to find additional significant characteristics that relate to anxiety. This will help them to standardise the platform to enable more accurate recognition of stress and anxiety. Aspects including head and mouth movements, lip and cheek deformations, eye gazing, and blinking rates may fall under this category. The final findings will offer a comparative assessment of both objective stress measurements [6].

Ngumimi Karen Iyortsuun , Soo-Hyung Kim , Min Jhon , Hyung-Jeong Yang and Sudarshan Pant performed analysis of existing system.In this study, 33 articles on the diagnosis of schizophrenia, depression, anxiety, bipolar disorder, post-traumatic stress disorder (PTSD), anorexia nervosa, and attention deficit hyperactivity disorder (ADHD) were

retrieved from various search databases using the preferred reporting items for systematic reviews and meta-analysis (PRISMA) review methodology. These publications were chosen based on their use of machine learning and deep learning technologies, individually assessed, and their recommended methodologies were then classified into the various disorders included in this study. They outlined some challenges in DL and ML approaches for mental health diagnosis in this review. small sample sizes are prevalent in the mental health field due to the high expense of data collecting that requires human participation. Although many ML models may demonstrate resilience when trained on a limited sample size of data without sacrificing performance accuracy,When experimenting with DL models, extensive training data are often required since they allow researchers to comb parameter space while also allowing the model to generalize to avoid overfitting hazards. overfitting impacts model performance in deep learning. having access to meaningful, high-quality, large-scale data in the mental health sector is a significant challenge [9].

Noah Lewis, Armin Iraji Theo G. M. van Erp, Vince D ,primary concerns were the creation of neural networks for feature learning and the implementation of an adaptive control unit for the fusion phase. Mid-fusion with attention model consists of a multilayer feedforward network, an autoencoder, a bidirectional long short-term memory unit with attention as the features extractor, and a linear attention module for managing modality-specific effect. Important brain characteristics and genes/biological pathways linked to schizophrenia were discovered by post hoc feature analyses. The suggested methodology successfully combines multimodal neuroimaging and genetic data to forecast mental diseases. Understanding the underlying etiological mechanisms of the salient traits the model discovered could be improved by interpreting them [20].

Md Azher Uddin, Joolekha Bibi Joolee, and Kyung-Ah Sohn suggested a novel deep multi-modal framework for an automated depression evaluation that successfully makes use of facial and verbal clues. To be more precise, they first divide the audio and video data into predetermined chunks. These segments are then used as input for the Spatio-Temporal Networks, which collect both spatial and temporal features and provide the features with the greatest contributions heavier weights. In order to extract the facial dynamics by encoding the structural characteristics, a dynamic feature descriptor based on Volume Local Directional Structural Pattern (VLDSP) is presented. Then, to compile the segment-level features for audio and video data, we use the Temporal Attentive Pooling (TAP) method. The multi-

modal factorised bilinear pooling (MFB) approach is then used to successfully combine the multi-modal data[28].

Shaunak Inamdar, Rishikesh Chapekar , Shilpa Gite & Biswajeet Pradhansuggested a strategy that makes use of machine learning algorithms and natural language processing techniques to identify indicators of mental stress in social media posts. The authors have determined which machine learning model performs the best on the Dreaddit dataset by evaluating the classification performance of a few different models.It was superior than the SVM model with ELMo (Embeddings from Language Models). in terms of F1 score, precision, and recall. All things considered, the suggested approach has the potential to significantly lower mental health issues among the majority of social media users by recognizing symptoms and offering help and support to overcome them.[8]

An overview of the existing applications of artificial intelligence in support of precision psychiatry is given in survey by Matthew Squires, Xiaohui Tao , Soman Elangovan , Raj Gururajan , Xujuan Zhou , U Rajendra Acharya and Yuefeng Li. To support each stage of the therapy cycle, cutting-edge algorithms are being deployed. These systems may be able to spot those who are dealing with mental health issues, enabling them to get the help they require and receiving therapies that are specifically suited to their needs. The tremendous clinical variability seen within diagnoses of depression is also being highlighted by unsupervised learning approaches, which are dismantling the discrete diagnostic categories now in use. Artificial intelligence also offers the chance to abandon current approaches that are focused on group averages and move towards therapy prescription that is evidence-based. Their investigation, however, indicates that a number of barriers are now impeding the development of data-driven paradigms in healthcare.Current limitations of treatment response systems include small sample sizes and model validation. Additionally, small sample sizes increase the likelihood of model overftting. [10]

Keyue Yan, Tengyue Li, Joao Alexandre Lobo Marques , Juntao Gao and Simon James Fong,Examines and summarises a number of recent papers on multimodal machine learning for illness detection, and also identifies areas for further investigation. In addition to removing the need to train each individual learning modality, multimodal learning enables models to process data from a variety of various sources and increases the robustness of models when dealing with diverse data. A growing body of research has been done in recent years on how to gather data from various sources and create precise multimodal machine

learning models or deep learning models for medical diagnosis.They discuss that, in fields of bioinformatics and medicine research, multi-modal machine learning plays an increasingly important role. [11]

Guramritpal Singh Saggu, Keshav Gupta, K. V. Aryaused the PHQ-8 Questionnaire scoreto propose a three-stage multimodal machine learning framework framework named DepressNet for depression diagnosis. Bidirectional Long Short-Term Memory (BLSTM) layer network has been proposed And Extended Distress Analysis Interview Corpus (E-DAIC) dataset was used for the training and validation of the proposed method with the use of multiscale temporal features from audio, video, and text modality and attention mechanisms for fusion [12].

Junli Liu and Haoyuan Wanguses applies computational intelligence to educational psychology. The recognition model is established by physiological signals.Utilize a single or a collection of data to explore and analyze college students' psychological states as well as their emotional traits. This paradigm of emotional psychoanalysis is intended to be validated for its applicability. According to the research findings, the experiment's end model functions well, demonstrating the success of the study and its experimental reference value. The deep learning-based approach put forth in this work has an altered application effect due to optimized computational intelligence. The research does still have several issues and shortcomings, though.The quality and quantity of the data need to be improved, as the data set established in this paper is monotonous; The model's shorter simulation times necessitate increasing the number of training samples and decreasing the overfitting level; Better modifications should also be incorporated into the model structure; In order to better understand emotion, more emotions and psychological states should be made available for machine learning, as well as improved human-computer interaction [18].

Jetli Chung and Jason Teo included a total 30 research publications in this review. They group the gathered research articles into categories based on mental health issues such schizophrenia, bipolar disorder, anxiety and depression, posttraumatic stress disorder, and mental health issues among children. In discussing the results, they consider the difficulties and restrictions that machine learning researchers have encountered while studying mental health issues [13].

T. E. Ramya,S. Sindhupriya proposed an approach, in which Twitter user comments are collected and a convolution neural network is used to train it. The suggested method

effectively helps in analyzing the emotional and mental behavior of the particular individual from the tweets that has been posted by the user It can quickly predict the characteristics of a certain class and uses less time, which aids in predicting users' emotional states and mental states. Python is used to create mental health prediction models, while Spyder is used as the environmental setup. To give a somewhat accurate rate of predicting users' mental states from their tweets, the model is run on test data. The models used in the classification of text such as Support Vector Machine, decision tree , Convolutional Neural Network and K Nearest Neighbor effectively classifies the labels that belongs to particular class [14].

An overview of conventional methods, technological substitutes, sensing instruments, tools for altering behaviour, and feedback interfaces are provided in the review article by Kieran Woodward , Eiman Kanjo , David J. Brown , T. M. McGinnity , Becky Inkster, Donald J. Macintyre , and Athanasios Tsanas. The difficulties posed by these technologies are then examined, with data collecting, privacy, and battery life among the crucial concerns that must be properly taken into account for the successful implementation of mental health toolkits. Modern machine learning algorithms can use the data that these ubiquitous devices can collect to produce reliable clinical decision support systems for diagnosis and real-time delivery of improvements to mental health [15].

In order to evaluate a user's mental health, focuses on identifying a data-driven statistic called the "Happiness Factor." To classify a post as "depressed," several models were trained. Based on the content of a user's posts, their "Happiness Factor" was determined. This measurement lets you know how depressed a person is. The outcomes demonstrate how well the classifier identified the depression level. Additionally, a system is developed for recommending mental health resources to users on their social media interface based on their "Happiness Factor."Depression identification is complicated by the fact that social media data frequently lacks some sorts of information about the person, such as sociodemographic traits, hidden behaviors, and psychological, social, and cultural environments. Individual users' post volumes differ significantly, and postings that contain an excessive number of phrases deemed important to the discussion could introduce bias into the diagnosis of depression. The studies examined herein may have had limited predictive value because they solely used text data from social media to investigate machine learning approaches to depression detection [16].

Ricardo Flores proposed an AudiFace multimodal deep learning model that screens for depression using temporal facial data, audio, and transcripts. AudiFace integrates a number of transfer learning models that have already been trained as well as bidirectional LSTM and self-Attention to incorporate all three modalities. For thirteen of the fifteen different datasets, AudiFace earns the greatest F1 scores when measured against the most advanced models. General wellbeing questions' ability to screen for depression is significantly enhanced by AudiFace. Both in the unimodal and multimodal models, eye gazing emerged as the most useful of the temporal facial variables.This research was limited by the number of participants in the existing dataset. This research is also unfortunately limited by the lack of datasets containing videos with depression labels[17].

Muhammad Arif, Grigori Sidorov, Helena G´omez-Adorno, Alexander Gelbukhused, five prevalent mental illnesses—depression, anxiety, bipolar disorder, ADHD, and PTSD and classified using unstructured user data from the Reddit site. They trained multi-class models for classical machine learning, deep learning, and transfer learning to identify mental diseases in people. The public health system will gain from this endeavour since it will automate the detection process and notify the proper authorities when someone needs emergency care [19].

To efficiently combine visual data from several modalities, such as the face and context,Bing-Jhang Lin , Yi-Ting Lin , Chen-Chung Liu , Lue-En Lee , Chih-Yuan Chuang , An-Sheng Liu , Shu-Hui Hung, and Li-Chen Fu used a Cross-Modality Graph Convolutional Network (CMGCN). In order to achieve better model convergence for multi-task learning, such as emotion detection and depression estimate, we have developed task-aware objective functions. To further transfer existing understanding of emotion to the depression model, They used the association between depression and emotion to build the Emotion Passer module. They used characteristics of schizophrenia to identify the mental condition during the detection phase. They conducted a number of tests on various benchmark datasets, and the results demonstrate that the suggested learning framework greatly improves state-of-the-art (SOTA) approaches [22].

Hanen Karamti, Eatedal Alabdulkreem, Hedia Zardi and Abeer M. Mahmoud examine the probability of depression using data taken from female social media profiles. To build the embedding words, a fresh deep learning model based on the psycholinguistic vocabulary is created. They start by taking the features out of the data both before and after preprocessing. Second, the data is labelled using the Convolutional Neural Network (CNN) in order to

extract the remaining features. The proposed model was able to accurately forecast the possibility of depression based on the preceding two phases. A thorough comparison study is also provided for the suggested system's evaluation [32].

Table 2.1 Summary of Literature

Types of mental illness	**References**	**Techniques used**	**Evaluation indicator**
Depression	[1],[2],[3],[4],[7],[21],[23], [24],[25],[26],[27],[29],30],[31]	Deep learning, Machine learning, Deep multimodal framework	F1 Score, RME, RMS, Precision, Recall
Stress and Anxiety	[6],[9],[20],[28]	CNN/LSTM, Machine learning	F1 Score, precision, Recall
Bipolar disorder, Post Traumatic StressDisorder (PTSD)	[15]	Deep learning, Machine learning, Transfer Learning	Accuracy
Schizophrenia	[12],[18]	Machine Learning,Deep Learning, Cross-Modality Graph Convolutional Network (CMGCN	Accuracy

2.1 Challenges and Gaps in Multi-modal Mental Health Analysis

From Literature review following gaps are identified.

1. Mental health assessment is restricted to subjective symptoms and observable signs.
2. Mental health status of the patient is continuously varying.
3. Mental disorder pose challenges due to symptom overlap and variability
4. To gather, process, fusion and analyze data from structured and unstructured sources is challenging.

Chapter 3 Objectives of Multimodal Mental Health Analysis Model

Fusing text, audio and video in a multimodal model involves combining information from three modalities to achieve a more comprehensive understanding of the data. Following are the objectives of proposed mental health analysis model.

1. To design robust fusion model to predict mental disorder using subjective and objective assessment.
2. To develop a model for analyzing and identifying risk factor for mental disorder.
3. To build a model to investigate overlapping of mental disorder.
4. To create dataset from video resources.

Chapter 4 Methodology for Mental Health Analysis Model

Fusion of text, audio, and video in multimodal models for mental health analysis is a complex and promising area of research with the potential to provide valuable insights and support for mental health professionals. This fusion allows for a more comprehensive understanding of an individual's mental state, as it considers multiple modalities of communication and expression.

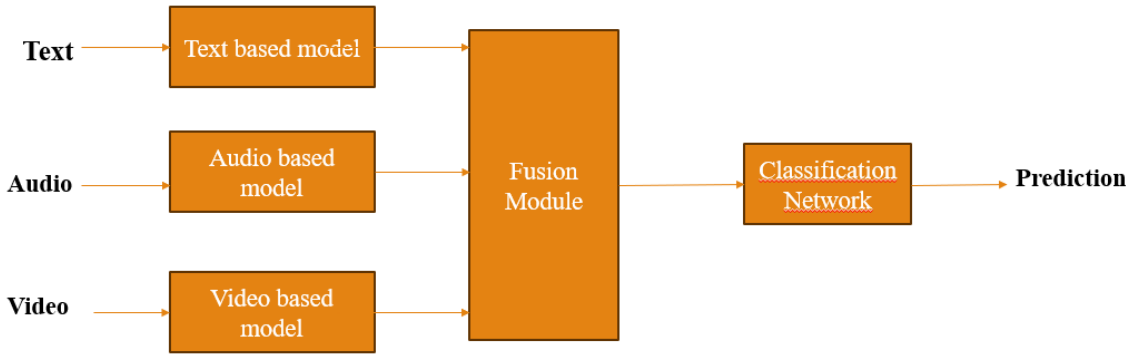

Figure 4.1 Proposed Multimodal Mental Health Analysis Model

Individual processing of each modality is known as encoding. In our proposed model three encoder are there, one for text, second for audio and third for video. The encoder extracts feature from the input data in each modality and converts them into a common representation that can be processed by subsequent layers in the model.

Text-based model can perform sentiment analysis on text data to determine the emotional tone or sentiment expressed. This is useful in monitoring changes in a person's emotional state over time. Textual data can be analysed to detect symptoms of mental health conditions, such as mood swings, sleep disturbances, or changes in behaviour. Audio-based model can analyse speech patterns, prosody (intonation, rhythm, and pitch), and vocal features to recognize and classify emotions in spoken language. This can help assess an individual's emotional state, such as detecting signs of depression, anxiety, or stress. Video-based models can analyse facial expressions and gestures to recognize emotions, helping assess an

individual's emotional state. For example, they can detect signs of sadness, happiness, anger, or anxiety.Changes in facial expressions, posture, or movement can indicate stress or anxiety. Video models can monitor these visual cues and provide insights into an individual's stress levels.

A fusion module combines the features extracted from each input modality, during the encoding phase. Output of unimodal are combine using various fusion techniques, such as early fusion, late fusion, or hybrid fusion, to create a joint representation of the data. In our model output of Text,Audio and Video model is given as input to fusion model.

- Early fusion involves concatenating the raw data from different modalities into a single input vector and feeding it to the network.
- Late fusion, on the other hand, involves training separate networks for each modality and then combining their outputs at a later stage.
- Hybrid fusion combines elements of both early and late fusion to create a more flexible and adaptable model.

A classifier that accepts the fused data and makes predictions. In our model it will predict mental health status like depression, anxiety, post-traumatic stress disorder.

Chapter 5 Scheduling and Planning

The research work is expected to complete in 3years and is modelled in terms of six months periods. The PhD progress scheduling and the execution plan with expected outcomes of each semester is mentioned in below table.

Year 2	**Year 2**	**Year 3**	**Year 3**	**Year 4**	**Year 4**
Odd Sem	**Even Sem**	**Odd Sem**	**Even Sem**	**Odd Sem**	**Even Sem**
IPS	**APS**	**IPS**	**APS**	**IPS**	**APS**
To design robust fusion model to predict mental disorder using subjective assessment.	To design robust fusion model to predict mental disorder using subjective and objective assessment.	Analysis of Risk Factor for mental health status	To develop a model for analyzing and identifying risk factor for mental disorder.	To create dataset from video resources.	To build a model to investigate overlapping of mental disorder.
	Publication of Paper 1		Publication of Paper 2		Publication of Paper 3
				Copyright filling	Report writing

Table 5.1 Scheduling and planning

Note:

IPS: Internal Progress Seminar

APS: Annual Progress Seminar

Chapter 6 Conclusion

The fusion of text, audio, and video modalities for mental health analysis represents a promising frontier in the field of mental health care and research. This multimodal approach allows for a more comprehensive and nuanced understanding of individuals' mental well-being by considering their diverse forms of expression and communication. Combining text, audio, and video data provides a holistic view of an individual's mental state. Text reveals thoughts and written emotions, audio captures tone and sentiment in speech, and video reveals facial expressions, body language, and other non-verbal cues. By harnessing the power of multimodal data and sophisticated AI models, we can move closer to a future where mental health care is more personalized, effective, and accessible.

By analysing various aspects, the following conclusion has been reached.

- Multi-modal model can improve accuracy and performance of mental health diagnosis.
- The fusion of text, audio, and video modalities provides a holistic view of an individual's mental state.
- Multimodal models can aid in the early detection of mental health issues by analysing subtle cues that may not be apparent in a single modality.

References

[1] Ming Fang , Siyu Peng , Yujia Liang , Chih-Cheng Hung , Shuhua Liu," A multimodal fusion model with multi-level attention mechanism for depression detection" Biomedical signal processing and control ,Volume 82, April 2023, 104561

[2] Yuran Ru,Ning Geng,Li Li"Multi-Modal Depression Detection Based on High-Order Emotional Features" 5th Artificial Intelligence and Cloud Computing Conference (AICCC) (AICCC 2022), December 17–19, 2022, Osaka, Japan. ACM, New York, NY, USA

[3] Yutong Li Zhenyu Liu,Li Zhou Xiaoyan Yuan Zixuan Shangguan Xiping Hu, Bin Hu, "A facial depression recognition method based on hybrid multi-head cross attention network"Front. Neurosci., 24 May 2023 Sec. Visual Neuroscience Volume 17 - 2023

[4]HyewonYoo, Hayoung Oh" Depression detection model using multimodal deep learning" Preprints (www.preprints.org) 2023

[5] Donglei Song, Qiang Shen, Haotian Feng , Rui Song , Fausto Giunchiglia , Hao Xu, "Personalized Multi-task Attention for Multi-modal Mental Health Detection and Explanation" ,Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45),2023

[6] Ali Kargarandehkordi, Peter Washington," Computer Vision Estimation of Stress and Anxiety Using a Gamified Mobile-based Ecological Momentary Assessment and Deep Learning: Research Protocol" , medRxiv April 2023

[7] Vandanaa, Nikhil Marriwalab, Deepti Chaudharyb,"A hybrid model for depression detection using deep learning", Measurement:Sensors25(2023)100587

[8] Shaunak Inamdar, Rishikesh Chapekar, Shilpa Gite & Biswajeet Pradhan, "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing", Human-Centric Intelligent Systems volume 3, pages80–91 (2023)

[9] Ngumimi Karen Iyortsuun, Soo-Hyung Kim, Min Jhon , Hyung-Jeong Yang and Sudarshan Pant, "A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis", Deep Learning Approaches on Mental Health Diagnosis. Healthcare 2023

[10] Matthew Squires, Xiaohui Tao , Soman Elangovan , Raj Gururajan , Xujuan Zhou , U Rajendra Acharya and Yuefeng Li ,“Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment”, Squires et al. Brain Informatics (2023)

[11]Keyue Yan, Tengyue Li, Joao Alexandre Lobo Marques , Juntao Gao and Simon James Fong,” A review on multimodal machine learning in medical diagnostics” MBE, 20(5): 8708–8726 (2023)

[12] Guramritpal Singh Saggu, Keshav Gupta, K. V. Arya,“DepressNet: A Multimodal Hierarchical Attention Mechanism approach for Depression Detection” International journal of Engineering Science 2022 15(1) 24-32

[13] Jetli Chung and Jason Teo “Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges”, Hindawi Applied Computational Intelligence and Soft Computing Volume 2022, Article ID 9970363

[14] T. E. Ramya,S. Sindhupriya ,“An Effective Approach for Mental Health Prediction Using Machine Learning algorithm”International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org CCICS - 2022 Conference Proceedings

[15] Kieran Woodward , Eiman Kanjo , David J. Brown , T. M. McGinnity , Becky Inkster, Donald J. Macintyre , and Athanasios Tsanas ,“Beyond Mobile Apps: A Survey of Technologies for Mental Well-Being”, IEEE Transactions On Affective Computing, Vol. 13, No. 3, July-September 2022

[16] Smita Ghosh, Sneha Ghosh , Diptaraj Sen and Pramita Das,“Depression Detection Using Machine And Deep Learning Models To Assess Mental Health Of Social Media Users”, David C. Wyld et al. (Eds): MLDS, IBCOM, CSIA, NLAI - 2022 pp. 01-12, 2022. CS & IT - CSCP 2022

[17] Ricardo Flores,“AudiFace: Multimodal Deep Learning for Depression Screening”, Proceedings of Machine Learning Research 182:1–22, 2022

[18] Junli Liu1 and Haoyuan Wang,“Analysis of Educational Mental Health and Emotion Based on Deep Learning and Computational Intelligence Optimization”Front. Psychol., 17 June 2022 Sec. Health Psychology Volume 13 - 2022

[19] Muhammad Arif, Grigori Sidorov, Helena G´omez-Adorno, Alexander Gelbukh “Mental Illness Classification on Social Media Texts using Deep Learning and Transfer Learning” July 2022 DOI:10.48550/arXiv.2207.01012 2022

[20] Noah Lewis, Armin Iraji Theo G. M. van Erp, Vince D. Calhoun “Deep multimodal predictome for studying mental disorders” Wiley September 2022Human Brain Mapping 44(1) DOI:10.1002/hbm.26077

[21] Irune Zubiaga and Raquel Justo ,“Multimodal feature evaluation and fusion for emotional well-being monitorization”Pattern Recognition and Image Analysis: 10th Iberian Conference, IbPRIA 2022, Aveiro, Portugal, May 4–6, 2022, Proceedings May 2022 Pages 242–254

[22] Bing-Jhang Lin , Yi-Ting Lin , Chen-Chung Liu , Lue-En Lee , Chih-Yuan Chuang , An-Sheng Liu , Shu-Hui Hung, and Li-Chen Fu , “Mental Status Detection for Schizophrenia Patients via Deep Visual Perception” IEEE Journal Of Biomedical And Health Informatics, Vol. 26, No. 11, November 2022

[23] Tanya Nijhawa, Girija Attigeri and T. Ananthakrishna ,“Stress detection using natural language processing and machine learning over social interactions”, Nijhawan et al. Journal of Big Data (2022) 9:33

[24] Anu Priya , Shruti Garg , Neha Prerna Tigga, “Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms” Procedia Computer Science, Volumn 167 2020

[25] Tianruo Xu”Elder Depression Detection by Multimodal Means” BDP Education and Psychology 2022

[26] Haibo Xu,Xiang Wu ,Xin Liu,”A measurement method for mental health based on dynamic multimodal feature recognition”, PubMed 2022

[27] Ravi Prasad Thati ,Abhishek Singh Dhadwal, Praveen Kumar1 ,Sainaba P ,“A novelmulti-modal depression detection approach based on mobile crowd sensing and task-based mechanisms”, Springer 2022

[28] Md Azher Uddin, Joolekha Bibi Joolee, and Kyung-Ah Sohn ,“Deep Multi-Modal Network Based Automated Depression Severity Estimation” IEEE transaction on affective computing 2022

[29] Ping-Cheng Wei, Kunyu Peng, Alina Roitberg, Kailun Yang, Jiaming Zhang and Rainer Stiefelhagen ,“Multi-modal Depression Estimation based on Sub-attentional Fusion” ECCV 2022

[30] Wu, Pingping, Wang, Ruihao, Lin, Han, Zhang, Fanlong, Tu, Juan; Sun, Miao ,“Automatic depression recognition by intelligent speech signal processing A systematic survey”, CAAI Transactions on intelligent technology 2022

[31] Zhe Sage Chen, Prathamesh (Param) Kulkarni, Isaac R. Galatzer-Levy, Benedetta Bigio, Carla Nasca and Yu Zhang ,“Modern views of machine learning for precision psychiatry” open access article under the CC BY-NC-ND Patterns 3, November 11, 2022

[32] Hanen Karamti, Eatedal Alabdulkreem, Hedia Zardi and Abeer M. Mahmoud.”A Multimodal Deep Learning Approach for Identification of Severity of Reflective Depression”,Journal of Electrical Evgineering and Informatics 2022

[33] https://www.nimh.nih.gov/health/topics/caring-for-your-mental-health